

# **The PEN&PAD Data Entry System: From prototype to practical system**

John Kirby, Ph.D. and Alan L Rector, M.D., Ph.D.  
Medical Informatics Group, Department of Computer Science,  
University of Manchester, UK

*This paper describes some of the issues addressed in the transition of the PEN&PAD from prototype clinical workstation to practical data entry system for use by general practitioners in the UK. Background and motivation of the PEN&PAD and GALEN projects are presented before the operation of the PEN&PAD user interface is described. A number of issues which have arisen in the development of the PEN&PAD Data Entry System are discussed.*

## **INTRODUCTION**

The prototype PEN&PAD data entry and medical records system was developed and evaluated during the period from 1988 to 1993<sup>1,2</sup> as part of a programme of user centred design and investigations of the foundations of medical records<sup>3,4</sup>. It led to a widely evaluated prototype and a radical approach to data entry based on a compositional model of medical terminology<sup>5</sup>. The compositional approach to medical terminology was an important source of the GALEN project<sup>6</sup> which aims to produce a re-usable and application independent model capable of supporting a wide range of applications.

Since the initial description of PEN&PAD there have been a variety of systems using point and click approaches with proposed or actual links to compositional terminology systems<sup>7,8,9</sup>, and there is a growing interest in the use of compositional models of terminology throughout medicine<sup>10</sup>. However, few of these systems are yet in widespread and routine clinical use.

The goal of the PEN&PAD programme has been to produce a useful and usable interface for clinicians. Early on, it concluded that though the mechanics of interfaces in existing systems were indeed a problem, the more fundamental problem was the nature of the clinical data used in computerised patient record systems. Typically the data were derived from mono-hierarchical coding schemes such as Read 2 or ICD-CM which consist essentially of structured lists of large numbers of fixed compound expressions.

The basic defects of systems based on such schemes seemed insurmountable:

- Navigating to terms is slow and cumbersome because of the size and structure of large coding schemes.
- A list of fixed terms, however long, rarely captures the detailed content required for clinical care. Clinical care requires that patients be described rather than labelled.

The PEN&PAD user interface was based on a fully compositional approach in which complex descriptions of patients' problems were built up by combining elementary terms. Following extensive user evaluation some of the main conclusions were:

- A graphical approach to finding clinical topics is more acceptable than an extensive search through a hierarchical list.
- Compositional data entry forms provide a simple intuitive means for entering detailed clinical information.
- User customisation of the content of the forms is essential for acceptance.
- Generation of readable natural language expressions from the data entered.
- Automatic production of relevant codes — ICD, Read, etc. depending on the location — is a fundamental requirement.

Since 1994, the PEN&PAD system has been redeveloped to produce a data entry system for use in conjunction with an existing Electronic Health Care Record (EHCR) system. The PEN&PAD Data Entry System (PEN&PAD DES) will, in the first instance, operate in conjunction with the VISION EHCR system marketed by VAMP Health Limited, the leading supplier of computer systems to General Practitioners in the UK. At the same time it has been adapted to take account of ongoing development of the re-usable GALEN Common Reference Model and to contribute to it.

The transition from academic prototype to practical data entry system and the adaptation to a more general re-usable model of terminology has required changes in both design and scope of the system.

- The user interface has been modified to accommodate a much larger number of clinical topics in the Topic Selector and to provide improved navigation to levels of greater detail on data entry forms.
- The method of using the model of terminology was extended through a system of views to enable it to share a much larger, re-usable model being developed by the GALEN project<sup>11</sup>.
- A revised approach to constructing forms as a structured collection of concepts has been developed as the basis for development and user customization tools.
- The focus has shifted to the production of a data management system for entering qualitative descriptive clinical data into an existing host EHCR system. Functions such as prescribing, medication records, allergies and warnings, etc., are implemented by the host EHCR system.
- The data entry system has been modularized and separated from the medical record proper in order to co-exist with more conventional software and database systems rather than the strongly object oriented schemes which underpinned the original PEN&PAD medical record.

This paper describes those changes and the resulting system. Evaluation of the system in use will be presented in a subsequent paper.

## USER INTERFACE

Entering data using the PEN&PAD DES takes place in two stages: first a presenting complaint or clinical topic is selected using the topic selector; then further details may be entered on a data entry form.

### Selecting a presenting complaint

The starting point of the user's interaction with the PEN&PAD DES will typically be the Topic Selector - other shortcut mechanisms are also available. The central feature of the Topic Selector is (shown in Figure 1) is the graphic of the human body which allows the user to select part of the body which is the location of the patient's symptom or disease by pointing and clicking.

Symptoms and diseases are associated with body locations, so that when a location is selected only

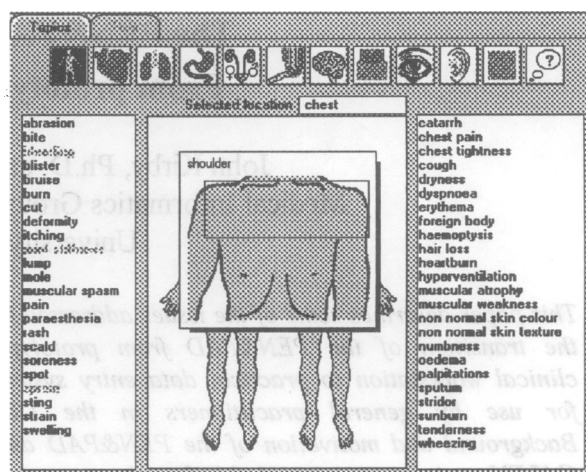


Figure 1: The Topic Selector

symptoms or diseases appropriate to that location are displayed in two lists on either side of the body graphic:

- The list on the left contains a list of presenting problems which may be located in many locations, e.g., pain, swelling, or fracture, allowing the selection of 'pain + chest', 'swelling + knee' or 'fracture + femur'. It is also possible to select a symptom or disease in more than one location, e.g., 'sunburn + head, neck and shoulder'.
- The list on the right contains those presenting complaints which are more specifically associated with a selected location, e.g., 'diarrhoea', 'palpitation' or 'cough'.
- General and systemic symptoms and diseases appear in the list on the right when "Whole body" is selected on the graphic.
- Selection of a body system icon at the top of the screen facilitates access to diseases of systems such as haematological or lymphatic by filtering both left and right hand lists to leave only those items relevant to the selected system.

### Compositional Data Entry Forms

A typical data entry form is shown in Figure 2 and consists of the following sections:

*Descriptors:* allows the user to enter information on such things as the Severity, Progress, Onset, Aggravating and Relieving factors of the topic of the form.

*Associated Symptoms:* allows the absence, presence or more details to be recorded on other clinically relevant symptoms.

**Figure 2: A data entry form**

**On Examination:** allows the user to enter clinically relevant findings from an examination of the patient.

**Diagnosis:** allows a diagnosis to be selected from a short list of diagnoses commonly associated with the presenting symptom - only relevant when the topic of the form is a symptom.

Some of the issues around how data entry forms are constructed will be discussed on the next section.

A major challenge in designing the operation of forms has been the need to provide initially brief forms with typical but limited detail while at the same time providing access to much greater detail should that be required. Three mechanism, corresponding to buttons shown in Figure 2, have been used:

**Additional:** provides access to other less commonly descriptors such as aggravating and relieving factors.

**More:** produces a pop-up form for entering greater details on an associated symptom or an examination finding.

**Further:** in the associated symptoms section allows data to be entered on any symptom available in the system; and in the examinations section allows data to be entered on any available examination.

Using these mechanisms the record shown in Figure 3, based on the selection shown in Figure 2, can be produced.

Severe, productive cough for 2 weeks unchanging, aggravated by exercise

Dyspnoea present

Rapid onset, mild haemoptysis for 2 days

**On examination**

**Respiratory system**

Moderate amount, green, bloody sputum

Breathing with slight difficulty

Dull percussion in left base

Expiratory wheeze in left apex

Coarse crackle in left base

**Cardiovascular system**

Heart sounds normal

**Diagnosis**

Chronic bronchitis

**Figure 3: A typical clinical record**

## UNDERLYING TERMINOLOGY SERVICES

At the heart of the PEN&PAD Data Entry System is an enhanced and re-implemented version of the GALEN Terminology Server, known as the Terminology Services Module (TSM). Like the GALEN Terminology Server, it provides automatic classification of concepts for use in information retrieval, conversion to existing classification systems, and encapsulation of complex concepts in fixed length records for storage in conventional relational databases. In addition, it provides facilities in two areas:

- Limiting the data input forms to what is sensible
- Grouping concepts pragmatically to match doctors' usual practice

Both sets of facilities are supported by authoring tools to tailor the data entry forms to new situations.

### Limiting the data input forms to what is sensible

The GALEN Representation And Integration Language is used in the TSM to create a model of the taxonomies of medical concepts along with rules - known as sanctions - determining which concepts may be combined together to form new composite concepts. In the GALEN Terminology Server two levels of sanctions are used:

- grammatical - high level sanctions, e.g. stating that diseases may be located in body parts
- sensible - more concrete statements, e.g. stating that fractures may be located in bones

Within the model, the sanctions control the creation of new composite concepts such as *fracture which hasLocation femur* — i.e. a "fracture of the femur".

In the original PEN&PAD system, the sanctions had also been used directly to control the data entry forms. A form consisted of all those things which it would be 'sensible' to say about a concept. In practice, using the more general and re-usable GALEN model, many sanctions are not clinically relevant. There are three common reasons for irrelevant sanctions. Sanctions may be:

- inappropriately abstract — *e.g.* all hollow bodies can define a space, but this is rarely relevant to data entry.
- used only for the construction of complex concepts — *e.g.* all objects may have a surface, so that we may speak about "the surface of the hand" etc. — but this is rarely relevant to a clinical data entry form.
- incidental inverse sanctions — *e.g.* viruses may infect the liver causing hepatitis, but the fact that the liver may be infected by a virus is not directly relevant to its examination.

These problems have been solved by introducing a new level of constraints known as "visibilities" which filter out unwanted sanctions and structure the output into a simple descriptor plus values format. A new level of request to the TSM now returns all and only those descriptors and values of that concept that are sensible and visible — *i.e.* those descriptors which are *clinically relevant* and conveniently packaged.

### Defining data entry forms

There are two aspects to authoring data entry forms for a given application:

- refining the individual concepts so that they are relevant to the context of the form; and
- organising the collection of individual concepts which make up a form.

### Refining the individual concepts

The 'raw' output from the TSM after the application of visibilities includes all of the clinically relevant descriptors for a concept with all of their potential values. Not all of these descriptors will be universally applicable. The relative importance and usefulness of different descriptors and values may vary greatly between clinical specialties. Even within a single specialty, doctors may differ in what they consider to be important. For these reasons a further level of processing is used in the construction of data entry forms. This allows the specification of: the order in which the descriptors appear on the form; and how the descriptors are 'layered' in the user interface.

Only the descriptors considered most relevant appear on the first 'layer' when a form is initially displayed. Other descriptors are hidden initially and revealed only when requested by the user. For example, all that a general practitioner records routinely about a cough are its duration, severity and productive character. These descriptors are therefore visible initially. However, in some cases more information needs to be recorded, in which case the next 'layer' of descriptors such as aggravating and relieving factors, sound character, and diurnal variation can be revealed by the click of an 'additional' button as described in the previous section.

### Collecting concepts together to create a form

Conceptually, a data entry form consists of a collection of concepts derived from the TSM grouped under headings such as presenting symptom, associated symptoms, examination, and common diagnoses. The grouping of items on a form is largely pragmatic based on knowledge of customary clinical practice. The list of commonly associated symptoms and the set of common diagnoses for a presenting complaint goes far beyond what can be called 'terminological'. It is therefore not part of the GRAIL model itself and must be represented separately.

For example, an examination of the hand might consist of observations about

- features: such as the size, colour and temperature of the hand
- functions: such as the range of movement or grip of the hand
- lesions: the absence, presence and, if present, a more detailed description of certain kinds of lesions which may occur in the hand

All of these categories of observations have an underlying connection to the hand as features, functions or locations, respectively. There are underlying GRAIL relationships between the lesion, sensation or function and the body location: *movement* which *isFunctionOf* hand and *hasRange limited*, *tenderness* which *actsOn* hand, and *lump* which *hasLocation* hand, etc.

However, in practice an examination of the hand, at least in General Practice, cannot include all possible functions, sensations, and lesions. Consequently for each symptom a small number of such concept is defined. For example, for symptom of pain in a given location, then tenderness is relevant; or for a

complaint of swelling, then the examination might include "generalised swelling" and "lump". In addition, there might be information from the model to refine the location and laterality.

To achieve adequate clinical focus for quick data entry, functions and lesions rather than locations have been made the focus of the examination, e.g. "pain in the hand", "swelling of the foot", "wheezes in the chest" etc. This approach has proved particularly effective as a means to develop brief relevant forms in functional areas such as the cardiovascular, respiratory, and nervous systems.

### USE WITH EXISTING EHCR SYSTEM

PEN&PAD DES provides a largely self-contained means of entering detailed structured clinical data into any existing EHCR system. The user invokes the PEN&PAD DES from within the host system. Optionally, patient data such as age, sex, etc., may be passed from the host system in order to provide a patient context for the data entry, e.g., avoiding the pregnant man or the woman with prostate problems.

When the doctor has accepted the entered data, the PEN&PAD DES returns to the host EHCR system: a TSM ID number; a natural language expression for the data entered; and corresponding ICD or Read code. How much of this information is stored in the host EHCR is determined by the application.

### CONCLUSION

The PEN&PAD DES combines the user interface designs of the original prototypes with a modified GALEN approach to terminology services in order to produce a practical data entry system designed to operate in conjunction with an existing EHCR system. To achieve a practical system has required re-implementing some features and enhancing others to meet the practical day to day requirements of convenience and speed. It is expected that the PEN&PAD DES will become widely available to GPs in the UK by the end of 1996.

### Acknowledgements

This work is supported in part by the Teaching Company Scheme of the UK Department of Trade and Industry and VAMP Health Limited.

### References

1. Nowlan, W. A. (1994). "Clinical workstation: Identifying clinical requirements and understanding clinical information." *International Journal of Bio-Medical Computing* 34: 85-94.
2. Horan, B., Rector, A., et al. (1990). Supporting a Humanly Impossible Task: The Clinical Human-Computer Environment. *Interact 90*, Elsevier Science Publishers, B.V. North-Holland. 247-252
3. Rector, A., Nowlan, W. and Kay, S. (1991). "Foundations for an Electronic Medical Record." *Methods of Information in Medicine* 30: 179-86.
4. Rector, A., Nowlan, W., et al. (1992). "A Framework for Modelling the Electronic Medical Record." *Methods of Information in Medicine* 32(2): 109-119.
5. Nowlan, W. and Rector, A. (1991). Medical Knowledge Representation and Predictive Data Entry. *Artificial Intelligence in Medicine Europe (AIME-91)*, Maastricht, Springer-Verlag. 105-116
6. Rector, A., Zanstra, P., Solomon, D. and The GALEN Consortium (1995). GALEN: Terminology Services for Clinical Information Systems. *Health in the New Communications Age* Ed. M. Laires, M. Ladeira and J. Christensen. Amsterdam, IOS Press. 90-100.
7. Lussier, Y. A., Maksud, M., et al. (1992). PureMD: A computerized patient record software for direct data entry by physicians using keyboard-free pen-based portable computer. *Sixteenth Annual Symposium on Computer Applications in Medical Care (SCAMC-92)*, Baltimore, MD, McGraw-Hill. 261-263
8. Moorman, P., van Ginneken, A., van der Lei, J. and van Bommel, J. (1994). "A model for structured data entry based on explicit descriptonal knowledge." *Methods of Information in Medicine* 33: 454-63.
9. Poon, A. D. and Fagan, L. M. (1994). PEN-Ivory: The design and evaluation of a pen-based system for structured data entry. *Eighteenth Annual Symposium on Computer Applications in Medical Care (SCAMC-94)*, Washington DC, Hanley & Belfus. 447-552
10. Cimino J. J., Clayton P. D., et. al. (1994). "Knowledge-based Approaches to the Maintenance of a Large Controlled Medical Terminology." *JAMIA*. 1(1), 35-50.
11. Rector, A., Solomon, W., Nowlan, W. and Rush, T. (1995). "A Terminology Server for Medical Language and Medical Information Systems." *Methods of Information in Medicine* 34: 147-157.